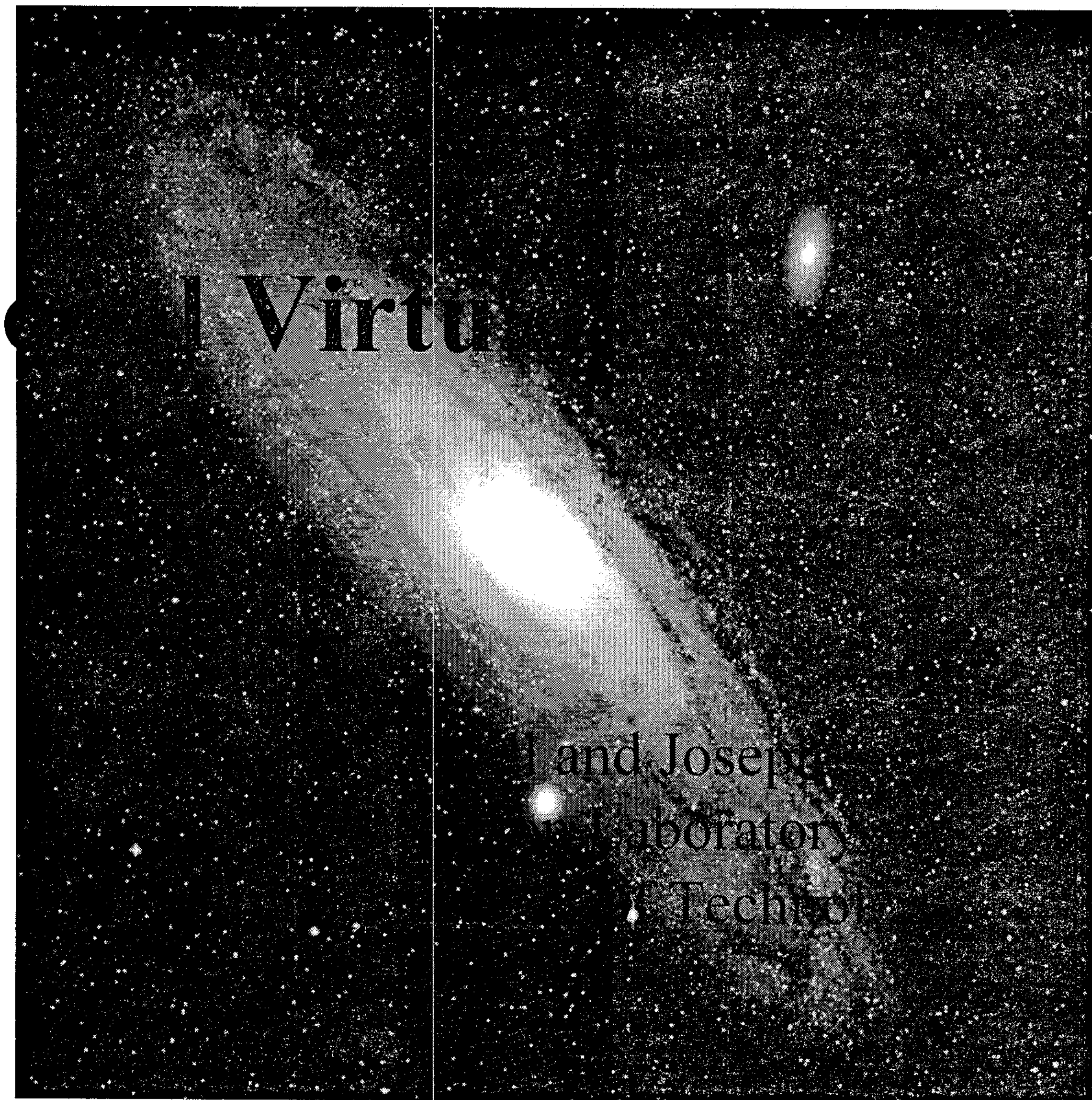
Nati… l Virtu… ry



December 4-5, 2001

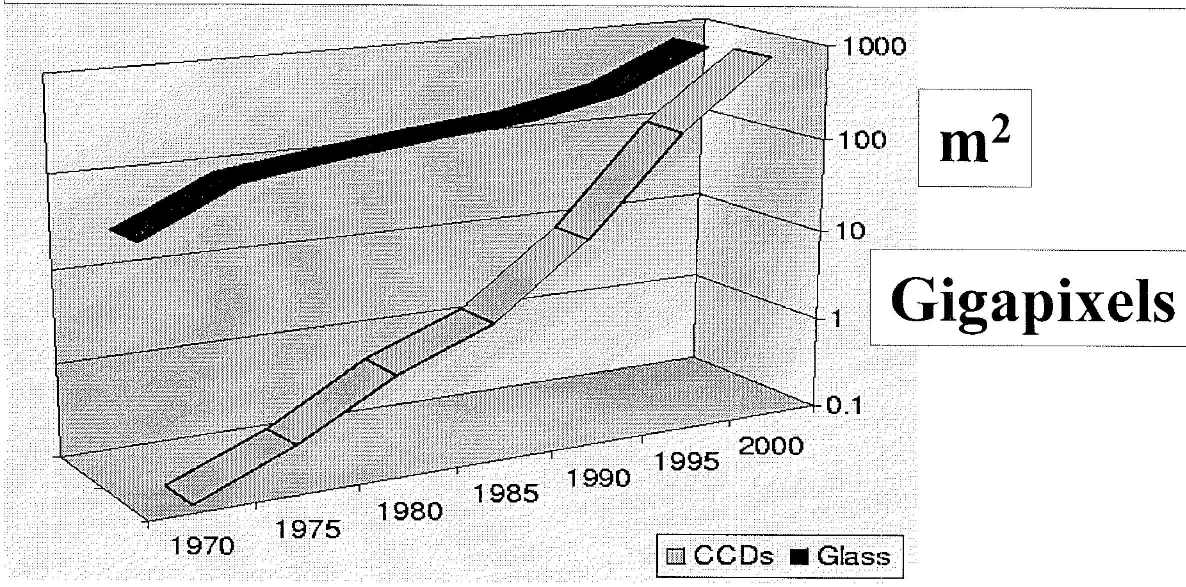
Space Science Applications of Information Technology Program
Earth Science Technology Office - Computational Technologies Project

Outline

- Overview of the NVO
- Mapping the NVO to the IPG
- NVO But a Single Example of Info Based Science Earth Science, Planetary Science, etc.
- The NVO Compute Node for Desktop Access
- yourSky Custom Mosaic Server
- Current & Planned IPG Activities

The Sky on Your Desktop

Growth in Aperture & Focal Plane Of Institutionally Managed Observatories



The NVO is conceptualized to deal with this avalanche of data, both image and catalog, already here and to arrive in the future

To meet these needs, a layered, interoperable software architecture needs to be wedded to the Internet II with its national network of computational resources.

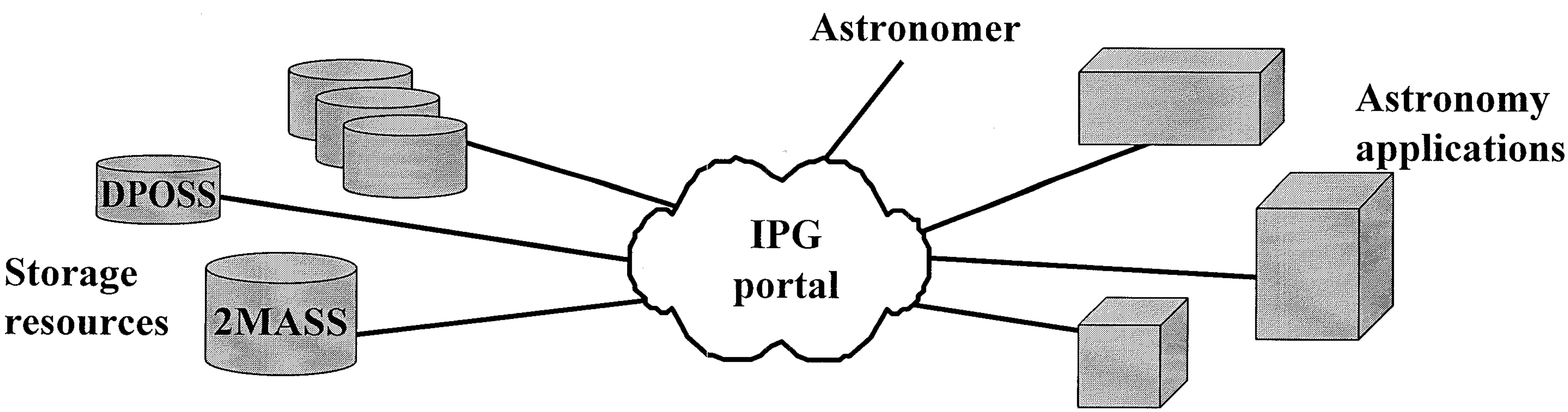
Goal:Deliver the data and its analysis to the individual astronomer's desktop

The National Virtual Observatory

- Large, widely distributed astronomy datasets: both object and image data
- High performance tools that need appeal to supercomputers
- Astronomers have limited computational, visualization and storage local to their desktop (circumscribes what they think they can do)

IPG and NVO a Natural Match

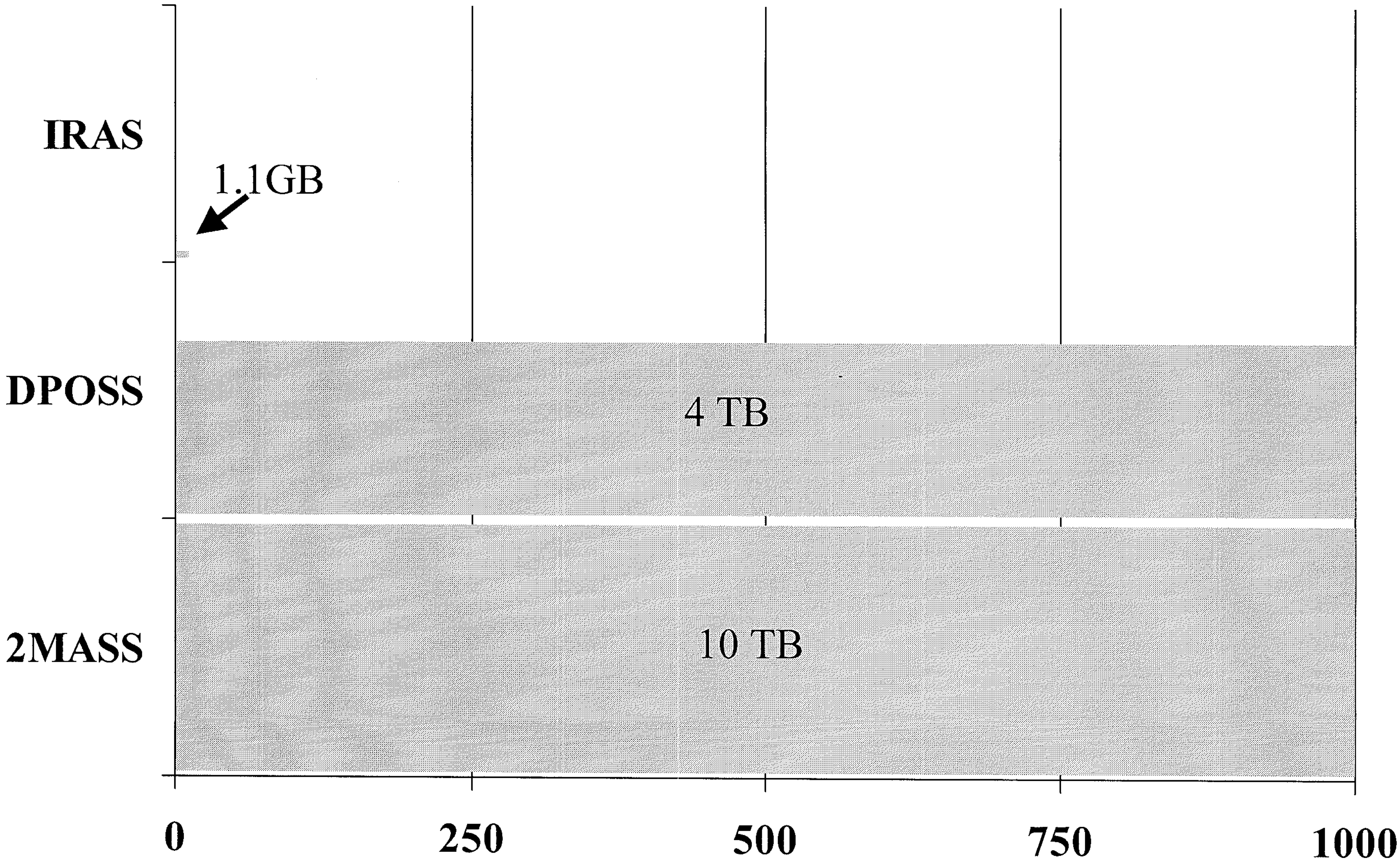
- Provide NVO astronomers access to the large astronomy datasets and high power processing of the data.
- Provide the middleware glue for a nationally distributed collection of image, object archives, computational power and high speed interlinks
- Help motivate & drive the IPG data intensive capability



Main Message

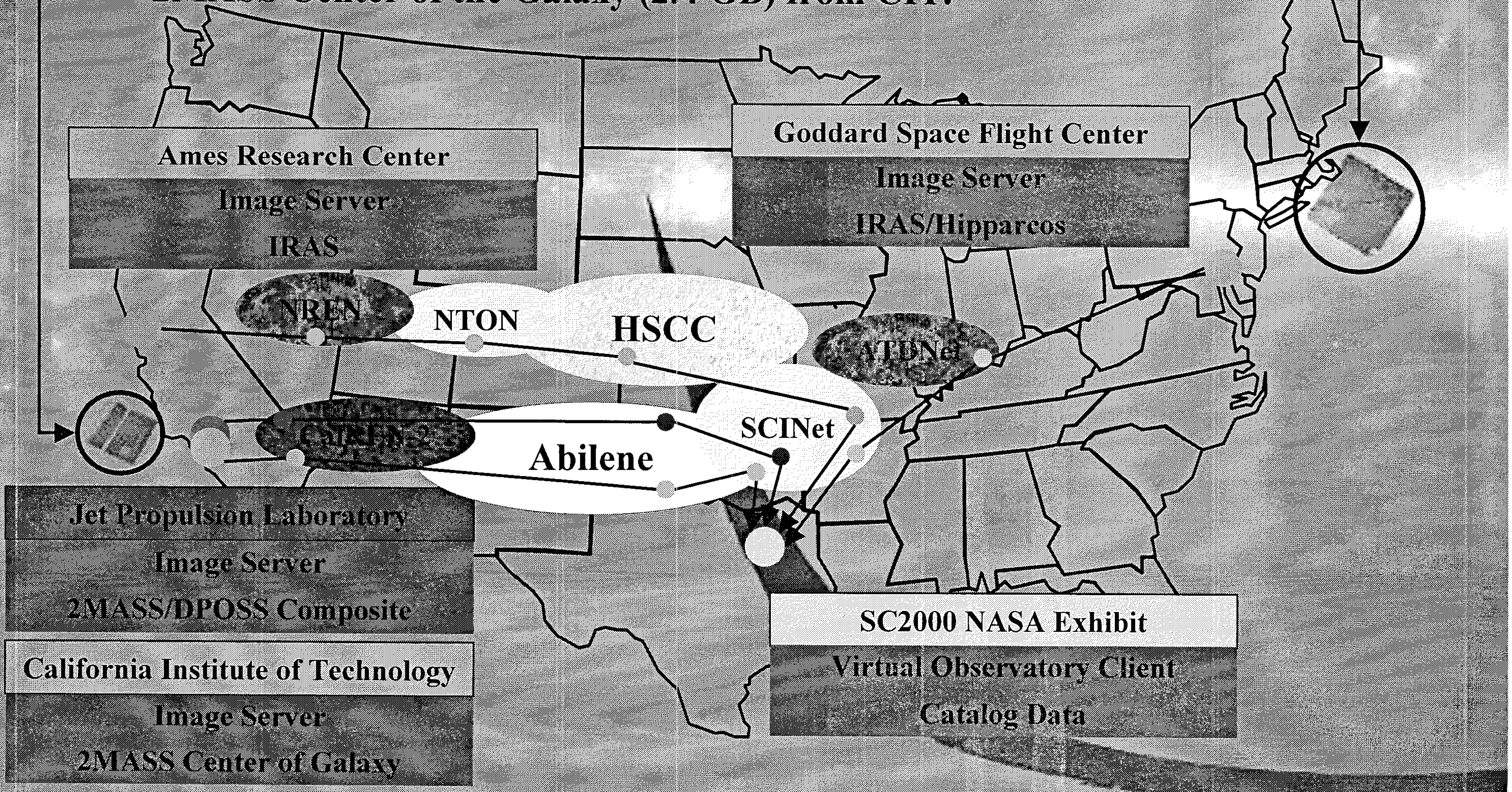
- The NVO will be community built, with emphasis on many, interoperable components, existing in a distributed high-performance, scalable environment.
- The clients can be anywhere and do not, themselves, need specialized nor high performance capabilities (hardware **or** software).
- This effort can be viewed as a prototype for several of NASA's "*information based science" needs. E.g.:*
  - Bringing realtime access and analysis to EOS DIS
  - To the Planetary Data System
  - The analysis and use of Interferometric Synthetic Aperture Radar
    - *Real time processing*
    - *Large area geodetic processing*

Data Archives



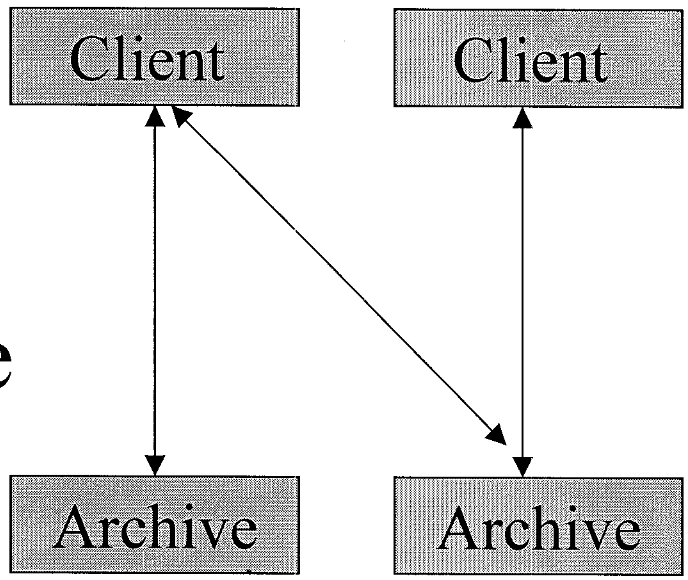
Digital Sky Virtual Observatory Demonstration at NASA Exhibit, SC2000, Dallas, TX, Nov. 4-10.

- IRAS (0.6 GB) served from GSFC and ARC.
- DPOSS/2MASS composite (10 GB) from JPL.
- 2MASS Center of the Galaxy (2.4 GB) from CIT.



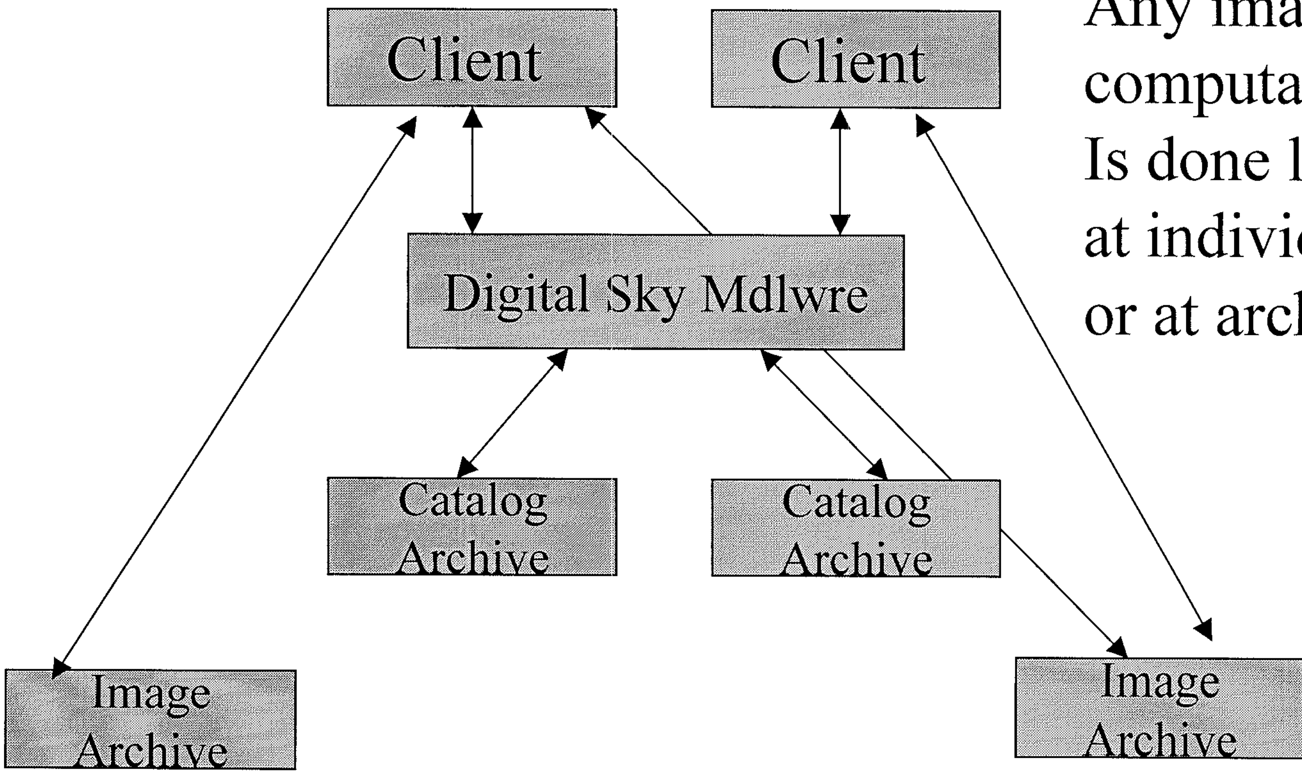
NVO Architecture Connects Clients to Data

Simple Architecture



But the NVO needs more power and flexibility than this

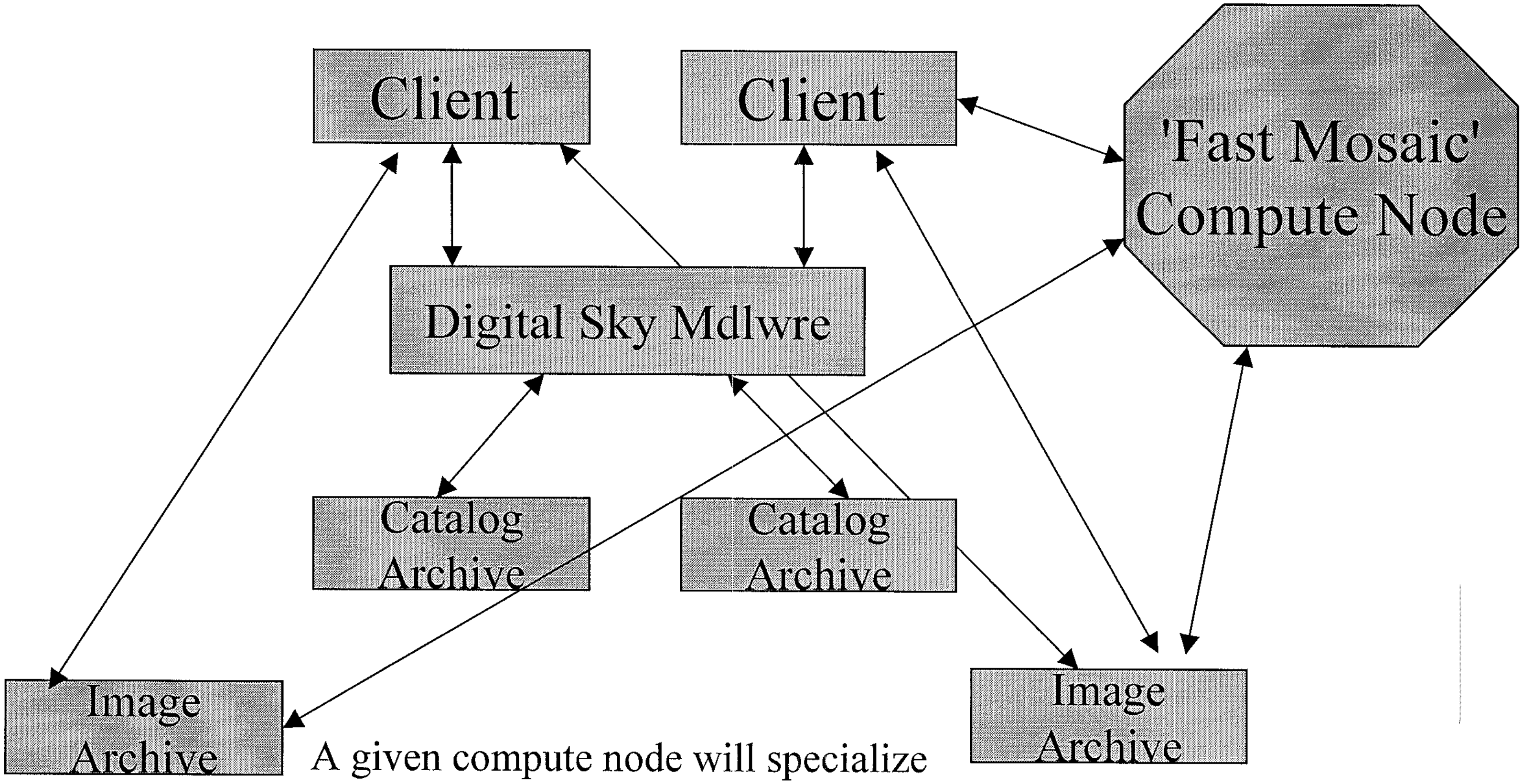
With Digital Sky



Any image analysis computation
Is done locally
at individual client
or at archive

The digital sky middleware provides
multi-archive catalog access and processing

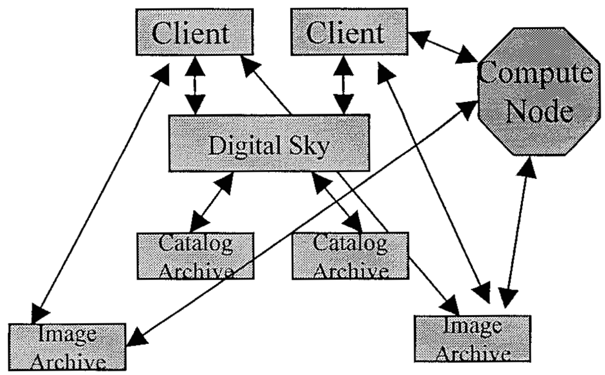
With Compute Node



A given compute node will specialize to a few services, organizing access to computational facilities (and itself behaving as a client to provide multi-archive access), performing the computations for its specialized functions, e.g. image processing, intensive catalog cluster processing ,etc.

Example: Custom 'On-the fly' Mosaic Node

- Help for composing and focusing a mosaic request
  - Pre created all-sky, all-survey browseable mosaics of medium resolution.
  - Ingestion of catalog search results plotted on the sky.
- Building a custom mosaic:
  - A sky region and survey set are selected
  - The node reaches into the 'original' image archives.
  - Searches for available Grid computing resources
  - Performs the mosaicking with custom: resolution, projection, image processing rules, re-projected catalog results, blending rules, etc.
  - Returns results to original client.



Principal Goal

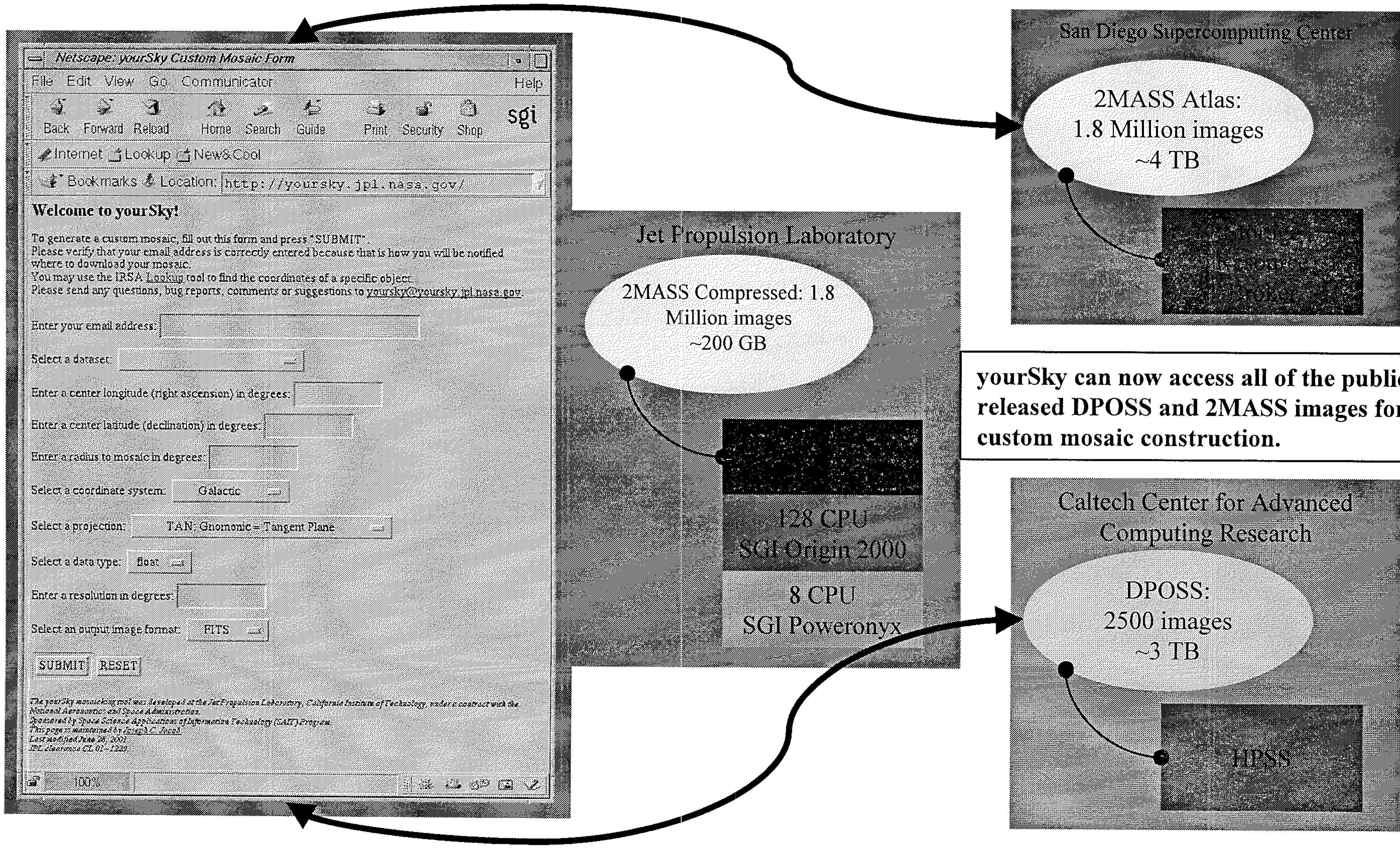
- **“Provide custom access to a compute-intensive, scalable interoperable service that delivers science-grade image mosaics to users desktops, through existing portals.”**
- Custom access = user defines size, sampling frequency, coordinates, projection, catalog entries
  - Architecture invites growth to expand options for custom image processing – multiple background removal techniques, overlap blending recipes, multiple surveys, etc.

Science Drivers for Mosaics

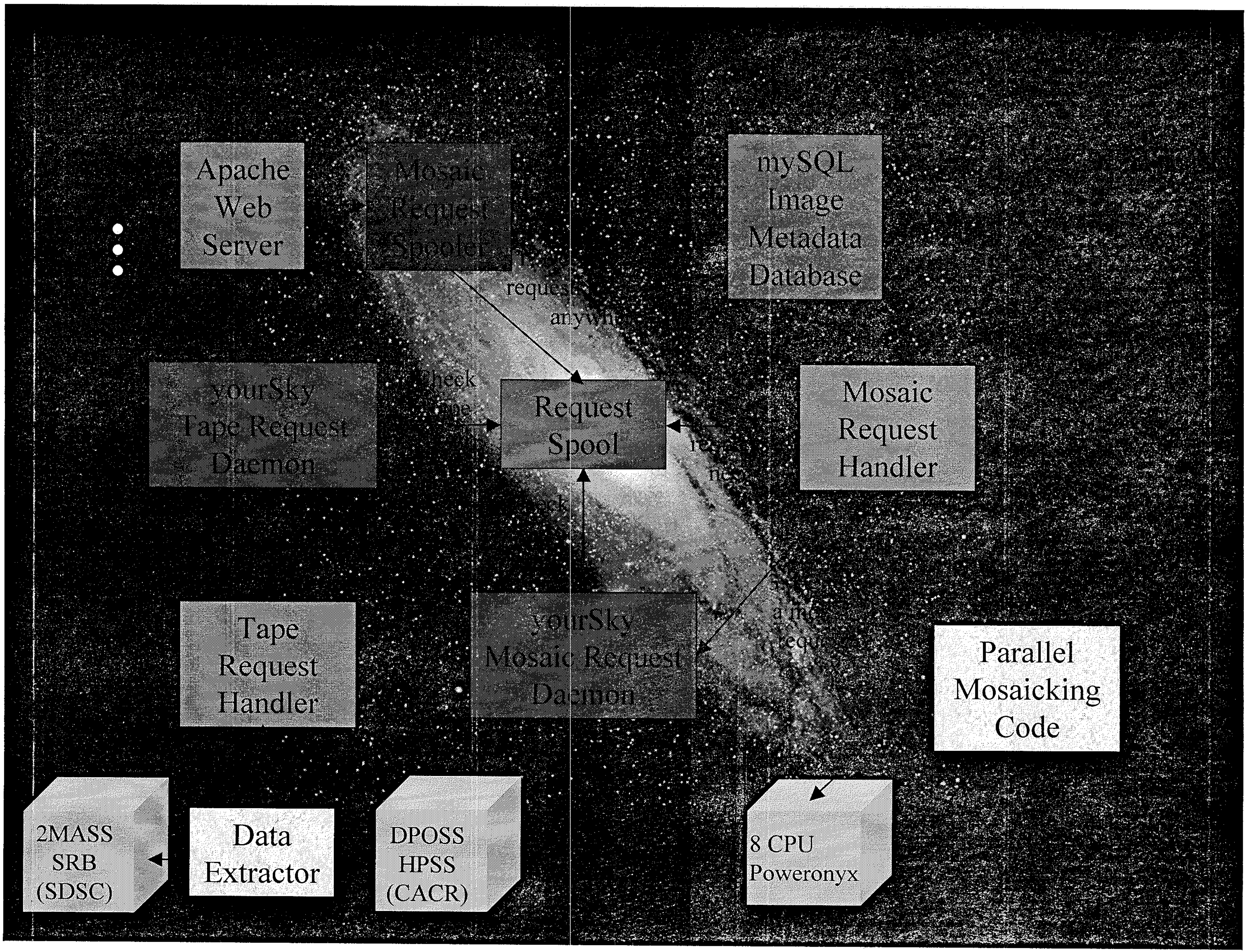
- Many important astrophysics questions involve studying regions that are at least a few degrees across.
  - Need high, uniform spatial resolution
  - BUT cameras give high resolution or wide area but not both => need mosaics
  - required for research and planning (e.g.SIRTF)
- Mosaics can reveal new structures & open new lines of research
- Star formation regions, clusters of galaxies must be studied on much larger scales to reveal structure and dynamics
- Mosaicking multiple surveys to the same grid – image federation – required to effectively search for faint, unusual objects, transients, or unknown objects with unusual spectrum.

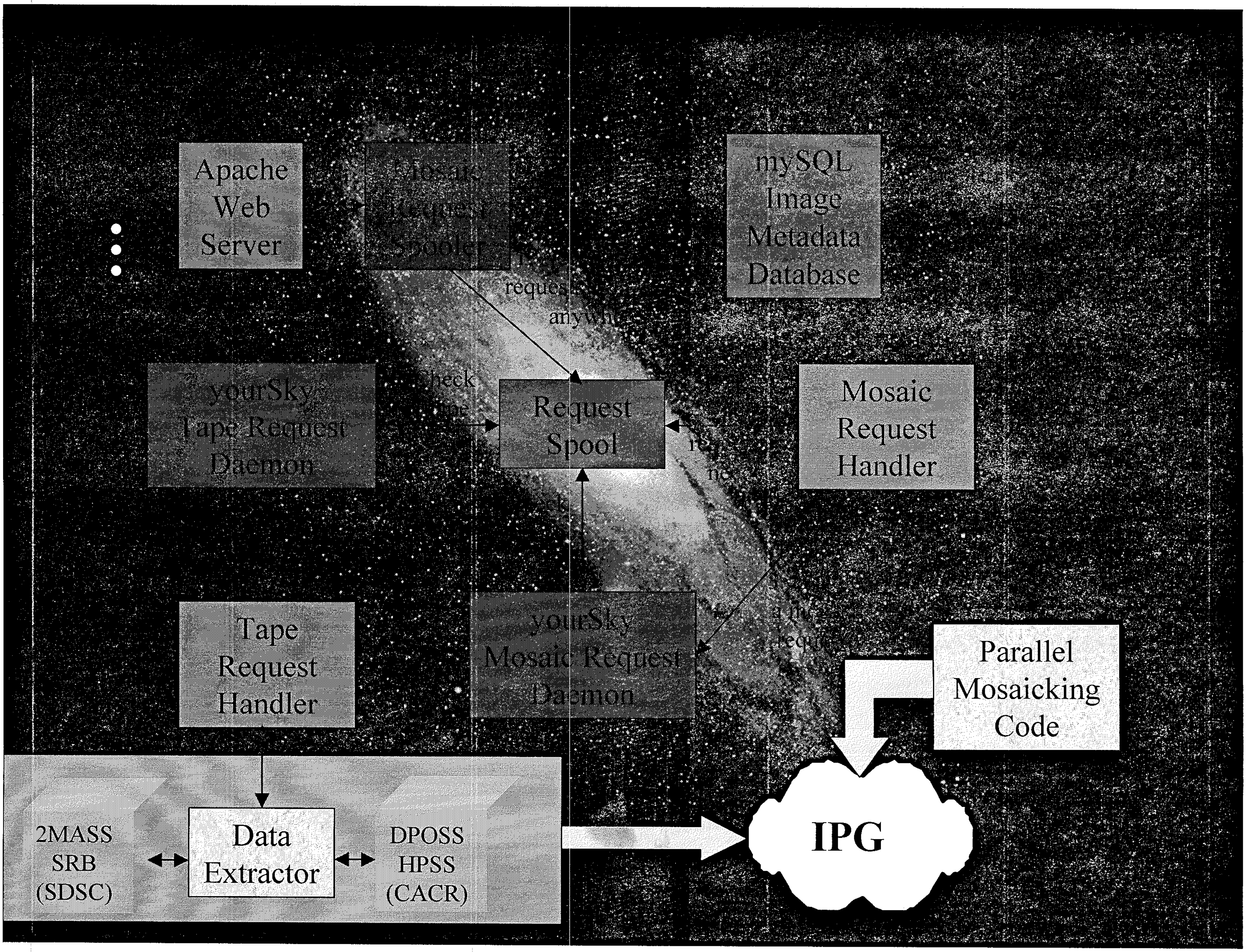
yourSky Custom Mosaic Server

http://yoursky.jpl.nasa.gov



yourSky can now access all of the publicly released DPOSS and 2MASS images for custom mosaic construction.


Apache Web Server
Mosaic Request Spooler
mySQL Image Metadata Database
yourSky Tape Request Daemon
Request Spool
Mosaic Request Handler
Tape Request Handler
yourSky Mosaic Request Daemon
Parallel Mosaicking Code
2MASS SRB (SDSC)
Data Extractor
DPOSS HPSS (CACR)
8 CPU Poweronyx



Apache Web Server
mySQL Image Metadata Database
yourSky Tape Request Daemon
Request Spool
Mosaic Request Handler
Tape Request Handler
yourSky Mosaic Request Daemon
Parallel Mosaicking Code
2MASS SRB (SDSC)
Data Extractor
DPOSS HPSS (CACR)
IPG

Parallel Image Mosaicking Code

For each input image, execute **Analyze** and **Build**.

Analyze:

- Examine input images to determine Right Ascension (longitude) and Declination (latitude) bounds, and pixel value ranges.
- Outer loop over **input** image pixels.

Build:

- Use coordinate system and projection transformations to map each output pixel back to the input pixel space.
- Sample from input with bilinear interpolation.
- Outer loop over **output** mosaic pixels.

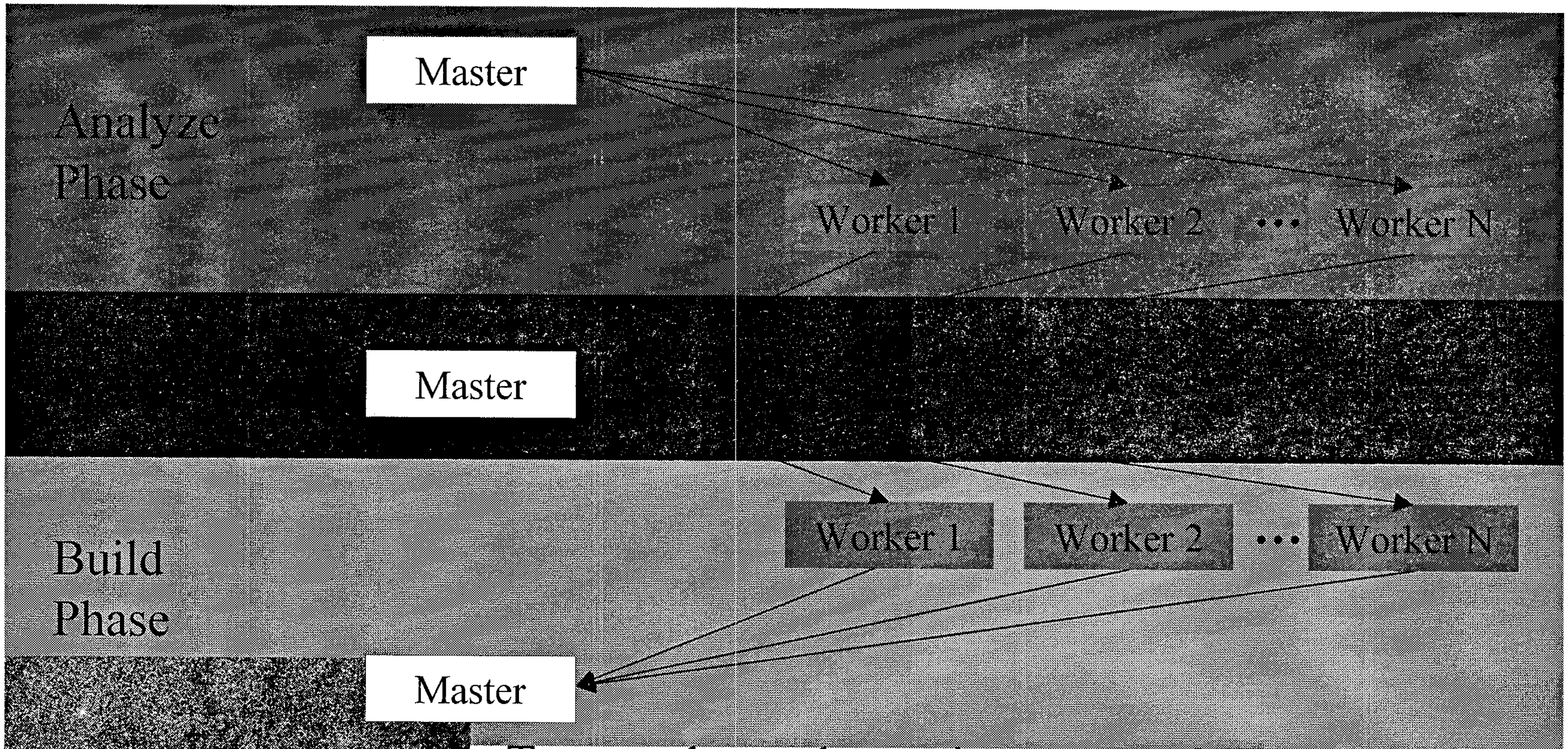
yourSky and the SRB

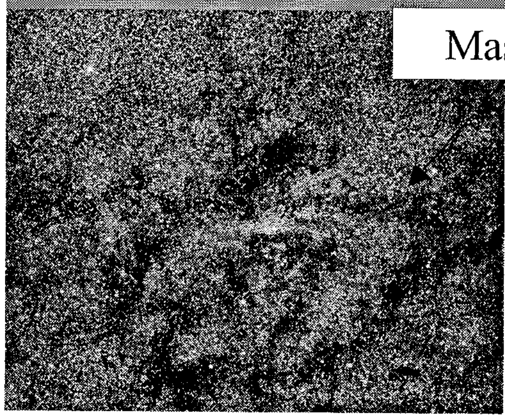
- Collaboration with George Kremenek, SDSC.
- Access to full release of 2MASS Atlas images in Storage Resource Broker (SRB) at SDSC.
  - Nearly 2 Million images (~ 40% of complete survey).
  - Each image ~ 2MB.
  - Total data size ~ 4 TB (~ 40% of complete survey).
- Two Versions:
  1. SRB client codes (Sinit, Scd, Sget, etc) used to retrieve the data to a local disk before the mosaic code is run.
     - Data is local by the time mosaic code is run, so computational resources do not have to be tied up while waiting for data retrieval from tape.
  2. SRB C API integrated with the mosaic I/O library to read image data directly from SRB at SDSC or NAS.
     - Requires no local disk to mirror data
     - Architecture supports "Run anywhere, Access SRB anywhere"

yourSky and Condor

- Collaboration with Peggy Li, JPL.
- yourSky mosaic code ported to Condor Master-Worker Model.
  - Data parallel application (embarrassingly parallel).
  - Well-suited to Master-Slave model
  - JPL has over 1500 Unix workstations
  - Typically 60% idle CPU cycles
  - Small condor pool of 14 Unix Workstations set up in two subnets at JPL.

yourSky and Condor, cont.



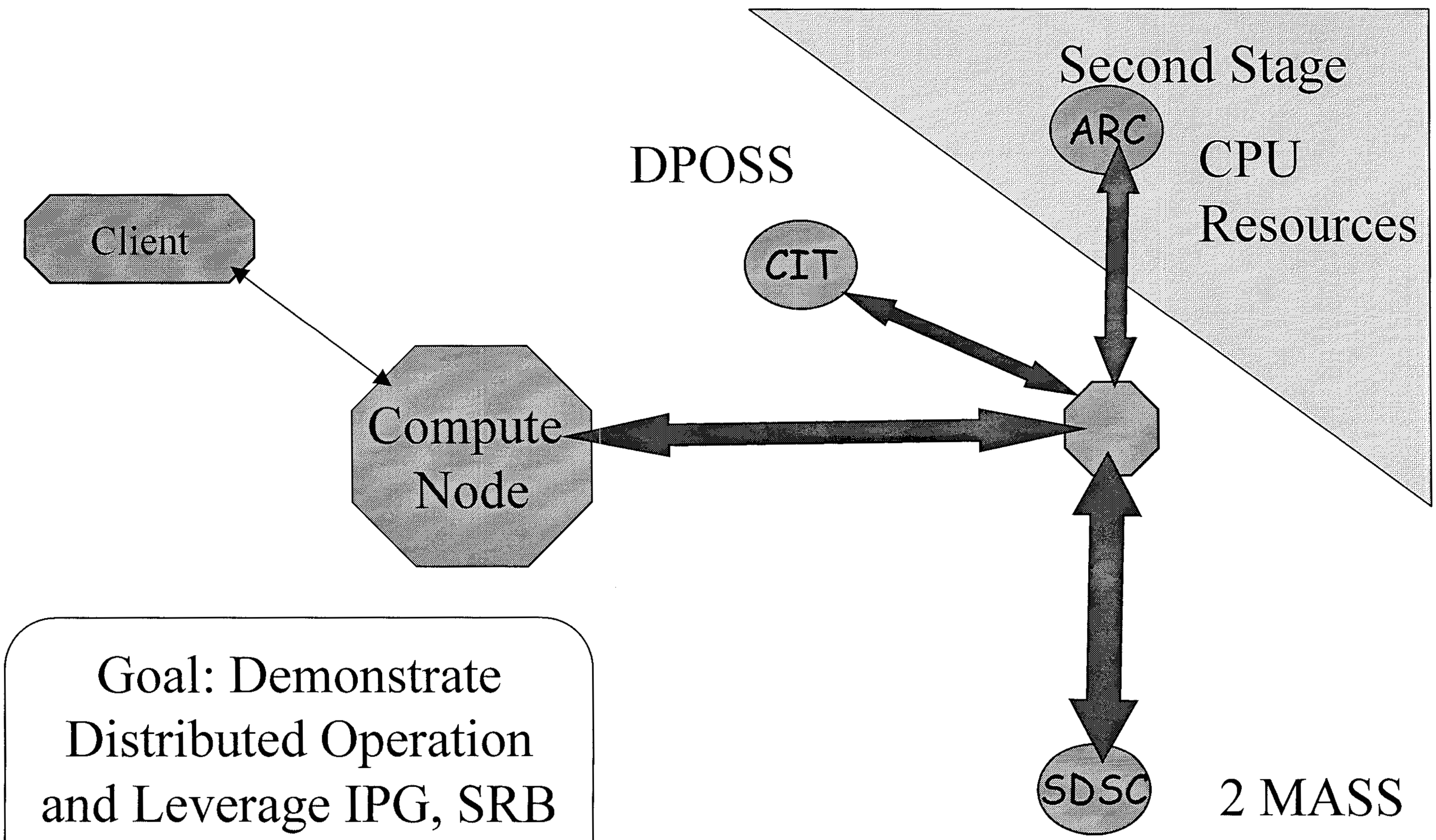


- Two worker tasks per image, one for Analyze and one for Build.
- Workers send single image results back to Master.
- Master combines the results into a single mosaic.

yourSky and Condor, cont.

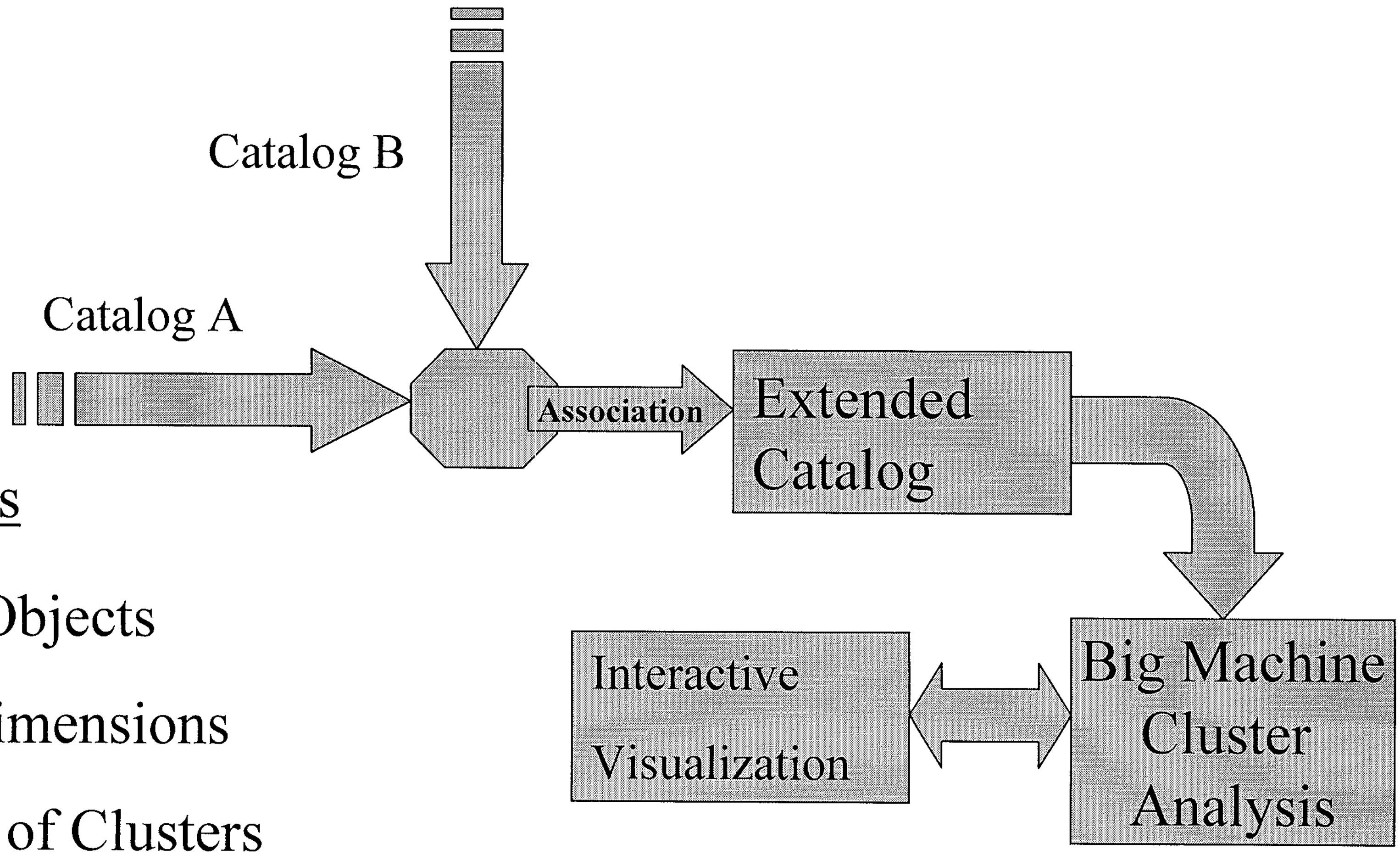
- Bottom Line Performance Comparison:
  - Efficiency for conventional run: 77.7%
  - Efficiency for Condor run: 50.4%
  - Difference represents Condor overhead and remote I/O delays.
- Conclusions:
  - Condor performance is acceptable.
  - At JPL, IT Security remains a concern.

Incorporate (ARC's) Information Power Grid and (SDSC's) Storage Resource Broker



Goal: Demonstrate Distributed Operation and Leverage IPG, SRB Infrastructure

Another Compute Node Might Specialize In Cluster Analysis



<u>Goals</u>

- 10^7 Objects
- 100 Dimensions
- Auto # of Clusters

Future Directions

- Goal: Use IPG Computational Resources with yourSky.
- Collaboration with Steve Klotz, ARC.
- Proposed Architecture:
  - ➢ JPL yourSky server receives mosaic request from user.
  - ➢ JPL yourSky Server retrieves 2MASS images from SDSC SRB and launches mosaic code via Globus on IPG machine at ARC.
  - ➢ CORBA Client/Server is used to send input images from yourSky server at JPL to Mosaic code running on IPG Machine at ARC.
  - ➢ When mosaic code completes, CORBA Client/Server is used to return mosaic result to yourSky Server at JPL.
  - ➢ JPL yourSky server makes result available to user.